

[subscribe](#) | [search](#) | [back issues on-line](#) | [order back issues](#)[ERCIM website quick index](#)**RESEARCH ACTIVITIES - ELECTRONIC PUBLISHING**

ERCIM News

No.20 - January 1995 - DRAL

Printing Hyperdocuments

by Simon Dobson and Victoria Burrill

Global hypertext systems such as the World Wide Web (WWW) are currently gaining in popularity, and may in future become the de facto means of delivering electronically a wide variety of information.

WWW technologies are designed to deliver information to a human reader, but do not as yet provide much support for machine-manipulable presentations of information. This makes tasks such as cross-server searching and document re-structuring more difficult: a good example is the problem of generating a paper copy of a hypertext document.

We have been investigating making WWW more amenable to automatic processing by importing ideas from the database world directly into HTML, WWW's document mark-up language. We have proposed a small extension to HTML which "publishes" the information structure of a set of pages as a database conceptual model. Elements of pages may be identified directly with entities and attributes from this underlying model, and relationships between entities may be described independently of the hypertext structure of the pages. We may then extract information from this "lightweight database" using database-like queries.

An important sample application of these techniques has been a system for "flattening" a set of hyper-pages when generating a (sequential) printed version of the hyperdocument. Our particular need is to generate from the Departmental web printed "hand-outs" of project descriptions for distribution to potential collaborators without Internet access. We mark-up the important elements of a project description (name, collaborators, aims, funding source etc) according to an entity-attribute-relationship model. We then formulate a query to extract the desired elements directly from the web. The query is template-driven, which allows extra formatting information to be added to the extracted information. This means that all printed hand-outs follow a common style despite variations in their hypertext descriptions.

Further information on our WWW activities may be found under: <http://web.inf.rl.ac.uk/proj/www/docs/>

PI ase contact:

Victoria Burrill - DRAL

Tel: +44 1235 44 6799

or Simon Dobson

Tel: +44 1235 44 5867

E-mail: {sd,vab}@inf.rl.ac.uk